Transient dynamics for sequence processing neural networks: effect of degree distributions

Yong Chen,[1, 2, ∗] Pan Zhang,[2, †] Lianchun Yu,[2] and Shengli Zhang[1]

[1]*Research Center for Science, Xi'an Jiaotong University, Xi'an 710049, China*
[2]*Institute of Theoretical Physics, Lanzhou University, Lanzhou 730000, China*
(Dated: February 1, 2008)

We derive an analytic evolution equation for overlap parameters, including the effect of degree distribution on the transient dynamics of sequence processing neural networks. In the special case of globally coupled networks, the precisely retrieved critical loading ratio $\alpha_c = N^{-1/2}$ is obtained, where N is the network size. In the presence of random networks, our theoretical predictions agree quantitatively with the numerical experiments for delta, binomial, and power-law degree distributions.

PACS numbers: 87.10.+e, 89.75.Fb, 87.18.Sn, 02.50.-r

I. INTRODUCTION

Recently, structure and dynamics in complex networks have attracted considerable attention and have been investigated in a large variety of research fields [1]. In particular, an important topic is whether the structure of neural wiring is related to brain functions.

Starting from pioneering milestone works that modeled Ising spin for neural networks, a large body of research has made a significant contribution to our understanding of parallel information processing in nervous tissue [2, 3, 4]. The equilibrium properties of the Hopfield model in a fully connected topology with the typical Hebbian prescription for the interaction strengths have been successfully described by a replica method [3, 4]. The dynamics of the fully connected Hopfield model with static patterns and sequence patterns have been widely studied using generating functional analysis [5, 6] and signal-to-noise analysis [7, 8, 9].

In recent years, there have been a large number of numerical studies of the Hopfield model on the complex structure, focusing on how the topology of a network, the degree distribution in particular, affects the computational performance of the formation of associative memories [10, 11, 12, 13, 14]. Various random diluted models have been studied, including the extreme diluted model [15, 16], the finite diluted model [17, 18], and the finite connection model [19]. However, to the best of our knowledge, here we derive for the first time the equation of retrieval dynamics for sequence processing neural networks with complex network topology.

In this paper, we study a modification of the Hopfield network, known as the sequence processing model, which acts as a temporary associative memory model [20, 21, 22, 23]. This model is very important to understand how the nervous system allows the learning of behavioral sequences because it requires hundreds of transitions that need to be precisely stored in neuronal connections [24]. The asymmetry of the interaction matrix rules out equilibrium statistical mechanical methods of analysis, including conventional replica theory. The goal of this work is to study the effect of the degree distribution on the transient dynamics of the sequence processing neural network. Using a probability approach, we derive an analytic time evolution equation for the overlap parameter with an arbitrary degree distribution that is consistent with our extensive numerical simulation results.

This paper is organized as follows. In Section II we introduce the definition of the sequence processing neural networks. The time evolution equations of the order parameters for the effects of degree distribution are derived and discussed in Section III. Section IV contains the comparison of theoretical results with numerical simulations. Finally, Section V presents a summary and the concluding remarks.

II. MODEL DEFINITION

In this paper, we consider a general version of sequence processing neural networks with parallel dynamics. The model consists of N Ising spin neurons $s_i \in \{-1, 1\}$. If the neuron i is at exciting status we put $s_i = 1$; otherwise

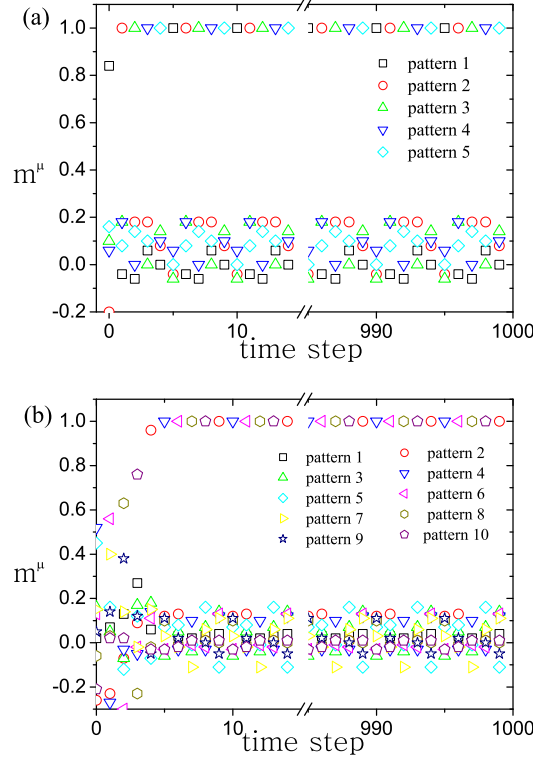∗Email address: ychen@lzu.edu.cn
†Email address: july.lzu@gmail.com

FIG. 1: (Color online) Temporal evolution of the macroscopic overlap parameters m^μ with respect to time t in the case of $w_{ij} = 1$. (a) $N = 100$, $p = 5$, and $\Delta p = 1$; (b) $N = 200$, $p = 10$, and $\Delta p = 2$.

(neuron i is inhibiting), we put $s_i = -1$. The embedded patterns are p states of the systems $\xi_i^\mu \in \{1, -1\}(\mu = 1, 2, \ldots, p; i = 1, 2, \ldots, N)$. The patterns are random so that each ξ_i^μ takes the values ± 1 with equal probability. The couplings between neurons are represented by the following form,

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu+\Delta p} w_{ij} \xi_j^\mu \qquad (\mu : \bmod\ p), \tag{1}$$

where $w_{ij} \in \{0, 1\}$ is a matrix element to tune the connection topology of coupling matrix $\mathbf{J}$ and $\Delta p = 0, 1, 2, \ldots, N-1$ describes the patterns learned as dynamic objects. This definition is clearly a typical asymmetric neural network. In the special case of $w_{ij} = 1$ and $\Delta p = 0$, the synapses are symmetric as in the Hopfield-Hebb networks.

The evolution dynamics of the systems are restricted to deterministic parallel dynamics where the spins are updated simultaneously according to

$$s_i(t+1) = \text{sgn}\left(\sum_j J_{ij} s_j(t)\right), \tag{2}$$

where $\text{sgn}(x)$ is the sign function. The time step is set to 1 in all our work.

In order to analyze the retrieval dynamics, the macroscopic overlap order parameter at time t is defined by

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu s_i(t) = \frac{2\lambda^\mu(t)}{N} - 1, \qquad \mu = 1, 2, \ldots, p. \tag{3}$$

Here, $\lambda^\mu(t)$ is the number of spins which have the same sign between the network state at time t and the given pattern ξ^μ. In Fig. 1, we present the evolutionary processes of the order parameter m^μ for the typical sequence

processing model with a global connection. It is found that the stored memories are reconstructed as a period-$(p/\Delta p)$ cycle: $2 \to 3 \to 4 \to 5 \to 1 \to 2 \to \cdots$ in Fig. 1(a), and $4 \to 6 \to 8 \to 10 \to 2 \to 4 \to \cdots$ in Fig. 1(b). To simplify the recent study, $\Delta p = 1$ in the following context. Note that the evolution equation for overlap parameters obtained with $\Delta p = 1$ is also suitable for any other Δp values.

III. TRANSIENT DYNAMICS AND MACROSCOPIC OBSERVABLE

Considering Eq. (1) and Eq. (2), one can get the following one-step update process,

$$
\begin{aligned}
s_i(t+1) &= \mathrm{sgn}\left(\sum_{j=1}^{N}\left(\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{\mu+1}w_{ij}\xi_j^{\mu}\right)s_j(t)\right) = \mathrm{sgn}\left(\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{\mu+1}\sum_{j=1}^{N}w_{ij}\xi_j^{\mu}s_j(t)\right)\\
&= \mathrm{sgn}\left(\frac{1}{N}\xi_i^{\nu+1}\sum_j w_{ij}\xi_j^{\nu}s_j(t) + \frac{1}{N}\sum_{\mu\neq\nu}^{p}\xi_i^{\mu+1}\sum_j w_{ij}\xi_j^{\mu}s_j(t)\right),
\end{aligned}
\tag{4}
$$

where ξ^{ν} is the νth stored pattern that is closest to state $\mathbf{s}(t)$ or $m^{\nu}(t) = \max\{m^1, m^2, \ldots, m^p\}$. The contributions for a single step evolutionary process in Eq. (4) consists of two parts denoted by

$$
h_i^1(t) = \frac{1}{N}\xi_i^{\nu+1}\sum_j w_{ij}\xi_j^{\nu}s_j(t),
\tag{5}
$$

$$
h_i^2(t) = \sum_{\mu=1,\mu\neq\nu}^{p}\xi_i^{\mu+1}\left(\frac{1}{N}\sum_j w_{ij}\xi_j^{\mu}s_j(t)\right).
\tag{6}
$$

The first part in the update function of Eq. (4), $h_i^1(t)$, drives the status of the ith spin to $\xi_i^{\nu+1}$ at time $t+1$. The other part, $h_i^2(t)$, is the noise term. In the case of absolutely stable and precise retrieving storage, $s_i(t+1) = \xi_i^{\nu+1}$. Since $s_i(t+1) = \mathrm{sgn}\left(h_i^1(t) + h_i^2(t)\right)$ and h_i^2 cannot change the sign of h_i^1. If $\xi_i^{\nu+1} = +1$, we have $h_i^1 > 0$. To ensure that $s_i(t+1) = \xi_i^{\nu+1}$, h_i^2 should satisfy $h_i^2 > -h_i^1$, so $h_i^2/h_i^1 > -1$. If $\xi_i^{\nu+1} = -1$, $h_i^1 < 0$, we also have $h_i^2/h_i^1 > -1$. So the probability of $s_i(t+1) = \xi_i^{\nu+1}$ is represented by the following equation,

$$
P\left(s_i(t+1) = \xi_i^{\nu+1}\right) = \sum_{z_i(t)=-1}^{(p-1)/m^{\nu}}P\left(z_i(t)\right), \quad z_i(t) = \frac{h_i^2(t)}{h_i^1(t)},
\tag{7}
$$

in which $P(z_i(t))$ is the probability of $z_i(t) = h_i^2(t)/h_i^1(t)$. Note that the degree of node i is $k_i = \sum_{j=1}^{N}w_{ij}$, which means that there are only k_i spins that are affected. For $\xi_i^{\nu+1} = 1$,

$$
h_i^1(t) = \frac{k_i}{N}m^{\nu}.
\tag{8}
$$

Applying the above equation to Eq. (7), we have

$$
P\left(s_i(t+1) = \xi_i^{\nu+1}\right) = \sum_{h_i^2=-m^{\nu}\frac{k_i}{N}}^{(p-1)\frac{k_i}{N}}P\left(h_i^2(t)\right), \quad (\xi_i^{\nu+1} = 1)
\tag{9}
$$

$$
= \sum_{h_i^2=-(p-1)k_i/N}^{m^{\nu}k_i/N}P\left(h_i^2(t)\right), \quad (\xi_i^{\nu+1} = -1).
\tag{10}
$$

With the following definition

$$
h_i'(t) = \sum_{\mu=1,\mu\neq\nu}^{p}\left(\frac{1}{N}\sum_j w_{ij}\xi_j^{\mu}s_j(t)\right).
\tag{11}
$$

We find that

$$\sum_{h'_i=-m^\nu \frac{k_i}{N}}^{(p-1)\frac{k_i}{N}} P(h'_i(t)) = \frac{1}{2}\left(\sum_{h_i^2=-m^\nu \frac{k_i}{N}}^{(p-1)\frac{k_i}{N}} P\left(h_i^2\right) + \sum_{h_i^2=-(p-1)\frac{k_i}{N}}^{m^\nu \frac{k_i}{N}} P\left(h_i^2\right)\right). \tag{12}$$

Combined with Eqs. (9-10), we have

$$P\left(s_i(t+1) = \xi_i^{\nu+1}\right) = \sum_{h'_i=-m^\nu k_i/N}^{(p-1)k_i/N} P(h'_i(t)). \tag{13}$$

Given the fact that the stored patterns are random and independent, the probability of the total number λ_N^μ spins that have has the same sign between $\mathbf{s}(t)$ and ξ^μ, $P(\lambda_N^\mu)$, is given by

$$P(\lambda_N^\mu) = C_N^{\lambda_N^\mu} 2^{-N}. \tag{14}$$

where $C_N^{\lambda_N^\mu}$ is used to denote a binomial coefficient, i.e. $C_N^{\lambda_N^\mu} = \frac{N!}{\lambda_N^\mu!(N-\lambda_N^\mu)!}$. For the precisely retrieved case (when system can retrieve patterns without error), $\mathbf{s}(t) \approx \xi^\nu$, equation (14) is correct obviously. Without the precisely retrieved condition, Eq. (14) in fact neglects the correlations between network states and pattern μ. If network topology has a local tree structure, the above deduction is correct. If there exist many short loops in the networks, the above deriving process is just an approximation. So, in this work, our derivations below are only suitable for sparsely connected random network, where the typical loop length is about $\log_{\bar{k}-1} N$ and $\bar{k}$ is average connection. In this paper, we use the condition $N \to \infty$, $\bar{k} \to \infty$, and $\bar{k}/N \to 0$.

Then from Eq. (3), we find

$$P\left(\frac{1}{N}\sum_j \xi_j^\mu s_j(t) = \frac{2\lambda_N^\mu}{N} - 1\right) = P\left(m^\mu = \frac{2\lambda_N^\mu}{N} - 1\right) = P(\lambda_N^\mu). \tag{15}$$

As noticed in the above statements [see Eqs. (4)-(6)], the local field is filtered by the topological structure of the networks. In this case, the probability of the total number $\lambda_{k_i}^\mu$ spins that have the same sign between $\mathbf{s}(t)$ and ξ^μ is

$$P\left(\frac{1}{N}\sum_j w_{ij}\xi_j^\mu s_j(t) = \frac{k_i}{N}m^\mu = \frac{k_i}{N}\left(\frac{2\lambda_{k_i}^\mu}{k_i} - 1\right)\right) = P\left(\lambda_{k_i}^\mu\right) = C_{k_i}^{\lambda_{k_i}^\mu} 2^{-k_i}. \tag{16}$$

Substituting into Eq. (10) the expressions of

$$x_i^\mu = \frac{2\lambda_{k_i}^\mu}{N} - \frac{k_i}{N} \tag{17}$$

we get the following form

$$P\left(\frac{1}{N}\sum_j w_{ij}\xi_j^\mu s_j(t) = x_i^\mu\right) = P\left(\lambda_{k_i}^\mu = \frac{k_i}{2}(1+x_i^\mu)\right) = C_{k_i}^{\frac{k_i}{2}(1+x_i^\mu)} 2^{-k_i}. \tag{18}$$

The other form of $h'_i(t)$ is readily deduced from the Eq. (17) which can be written as

$$h'_i(t) = \sum_{\mu\neq\nu}^p x_i^\mu = \frac{2\sum_{\mu\neq\nu}^p \lambda_{k_i}^\mu}{N} - (p-1)\frac{k_i}{N}. \tag{19}$$

Therefore, after coarse graining like in Eqs. (16, 18), using the definition of $\Lambda_{k_i} = \sum_{\mu\neq\nu}^p \lambda_{k_i}^\mu = \frac{Nh'_i}{2} + \frac{(p-1)k_i}{2}$, we obtain that

$$P_1 = P\left(s_i(t+1) = \xi_i^{\nu+1}\right) = \sum_{h'_i=-m^\nu k_i/N}^{(p-1)k_i/N} P(h'_i(t)) = \sum_{\Lambda_{k_i}=\frac{(p-1)}{2}k_i-\frac{m^\nu}{2}k_i}^{(p-1)k_i} P\left(\Lambda_{k_i} = \sum_{\mu\neq\nu}^p \lambda_{k_i}^\mu\right)$$

$$= \sum_{\Lambda_{k_i}=\frac{(p-1)}{2}k_i-\frac{m^\nu}{2}k_i}^{(p-1)k_i} C_{(p-1)k_i}^{\Lambda_{k_i}} 2^{-(p-1)k_i} = \sum_{\Lambda_{k_i}=0}^{\frac{(p-1)}{2}k_i+\frac{m^\nu}{2}k_i} C_{(p-1)k_i}^{\Lambda_{k_i}} 2^{-(p-1)k_i}. \tag{20}$$

By introducing the degree distribution of the network structure $P(k)$, where k is the number of links connected to a node, the total number $\lambda^{\nu+1}$ spins between the status $\mathbf{s}(t+1)$ and the stored pattern $\xi^{\nu+1}$ can be expressed as

$$\lambda^{\nu+1} = \sum_k N P_1 P(k). \tag{21}$$

Then from Eq. (3), we obtain the macroscopic observable

$$m^{\nu+1}(t+1) = \frac{1}{N}\sum_i \xi_i^{\nu+1} s_i(t+1) = \sum_k 2P_1 P(k) - 1. \tag{22}$$

Combining Eq. (20) and Eq. (22) and replacing Λ_{k_i} by n, we arrive at the evolution equation of the overlap parameter

$$m^{\nu+1}(t+1) = 2 \times \sum_k P(k) \frac{\sum_{n=0}^{\frac{1}{2}(p-1)k+\frac{1}{2}m^{\nu}(t)k} C_{(p-1)k}^n}{2^{(p-1)k}} - 1. \tag{23}$$

In the case of successful storage, the network finally tends to converge into a stable periodic cycle, or $m^{\nu}(t) = m^{\nu+1}(t+1)$. Finally, replacing $m^{\nu}(t) = m^{\nu+1}(t+1)$ by m_f, one has the iterative solution for the final overlap parameter,

$$m_f = 2 \times \sum_k P(k) \frac{\sum_{n=0}^{\frac{1}{2}(p-1)k+\frac{1}{2}m_f k} C_{(p-1)k}^n}{2^{(p-1)k}} - 1. \tag{24}$$

Note that it is difficult to calculate the above iterative equation of the overlap parameter at very large $(p-1)k$ due to the computational complexity of the factorial term. A reasonable solution is to replace the binomial distribution in Eq. (20) by a Gaussian distribution with the same expectation value and the same standard deviation. As a result, replacing n by x, we reformulate Eq. (23) as

$$
\begin{aligned}
m^{\nu+1}(t+1) &\approx \sum_k 2P(k) \sum_{x=0}^{\frac{k}{2}(p-1+m^{\nu}(t))} \sqrt{\frac{2}{\pi(p-1)k}} e^{\frac{-2\left(x-\frac{p-1}{2}k\right)^2}{(p-1)k}} - 1 \\
&= \sum_k 2P(k) \int_0^{\frac{k}{2}(p-1+m^{\nu}(t))} \sqrt{\frac{2}{\pi(p-1)k}} e^{\frac{-2\left(x-\frac{p-1}{2}k\right)^2}{(p-1)k}} dx - 1.
\end{aligned} \tag{25}
$$

Substituting into the above equation the following expression,

$$z = \frac{x-(p-1)k/2}{\sqrt{(p-1)k}/2}, \tag{26}$$

we find Eq. (25) under the form

$$
\begin{aligned}
m^{\nu+1}(t+1) &= \sum_k 2P(k) \sqrt{\frac{1}{2\pi}} \int_0^{\frac{m^{\nu}(t)}{\sqrt{(p-1)/k}}} e^{-z^2/2} dz \\
&= \sum_k P(k) \mathrm{erf}\left(\frac{m^{\nu}(t)}{\sqrt{(p-1)/k}}\right) \\
&\approx \sum_k P(k) \mathrm{erf}\left(\frac{m^{\nu}(t)}{\sqrt{p/k}}\right),
\end{aligned} \tag{27}
$$

where $\mathrm{erf}(\cdot)$ is the error integral function

$$\mathrm{erf}(z) = \sqrt{\frac{2}{\pi}} \int_0^z \exp(-x^2/2) dx.$$

Note that in the special case of fully connected networks, $\bar{k} = N$ and $P(k) = \delta(k-N)$. Herein, Eq. (27) is reduced to

$$m^{\nu+1}(t+1) = \mathrm{erf}\left(\frac{m^{\nu}(t)}{\sqrt{p/N}}\right). \tag{28}$$

which is a well-known result [7, 25]. It is interesting to compare the above equation with our result for arbitrary degree distributions. In Eq. (27), there is only a little modification. It should also be mentioned that Eq. (27) is also found for sparsely connected Hopfield networks [26].
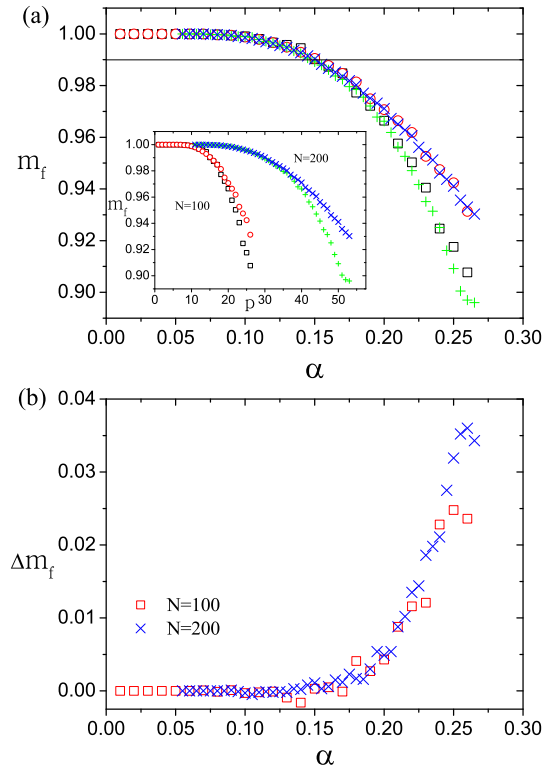
FIG. 2: (Color online) The final overlap m_f in the globally coupled network with $N = 100, 200$. (a) ($\circ, \times$): iterative results from Eq. (23), ($\square, +$): simulation results. The solid line is $m_f = 0.99$ which corresponds to $\alpha \approx 0.15$ from simulations and iterative results. The inset is the same result for abscissa p. (b) The iterative error Δm_f versus the loading ratio α. Each simulation point represents an average of 200 trials.

IV. NUMERICAL STUDIES

To verify the theory, we performed extensive simulations which are reported in this section. First of all, we give a succinct description of how our calculations for the iterative Eq. (24) were made. Note that a precondition of our work is that the network is capable of memorizing patterns in the form of its equilibria in each trial [see Eqs. (7, 21)]. The initial overlap m_f^0 in the right side of the iterative equation is set to 1.0. Then the second m_f^1 is obtained and is set as the initial overlap to calculate the third one. This process is repeated until $m_f^n \approx m_f^{n+1}$ within allowed precision. Thus, we arrive at the final stable macroscopic overlap parameter $m_f = m_f^n$ after n iterative steps. In our simulations, it is found that the iterative procedures always converge quite rapidly, stopping after $4-5$ steps at most.

In order to compare the different effects of various degree distributions on the performance of neural networks, we consider the following two cases. One case is the globally coupled network. Although this case appears somewhat trivial, it is helpful to compare our study with some common conclusions. The other case is the random network with various degree distributions, including the delta function, binomial, and power-law distributions.

A. Globally coupled networks

In this case, all the neuronal spins are connected with each other at any time, namely $w_{ij} = 1$, and the degree distribution $P(k)$ is a δ-function. This actually introduces a huge waste of energy but provides a neural network with the maximal retrieving performance. A large number of numerical and theoretical studies on this network dynamics have been made in the last two decades. These studies revealed that there exists a critical loading ratio known as the Amit-Gutfreund-Sompolinsky (AGS) value $\alpha_1 \approx 0.139$ for symmetric networks [3], a saturated stored capacity
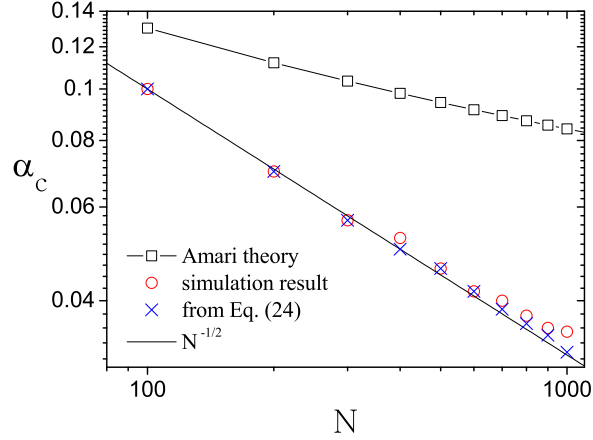
FIG. 3: (Color online) The critical precise loading ratio $\alpha_c = p_c/N$ for globally coupled networks. The solid line is $\alpha_c = 1/\sqrt{N}$. (∘): Simulation results. Each point represents an average of 200 trials. (×): Iterative results from Eq. (24) with precision $\Delta m \le 10^{-4}$. (□): results from the Amari theory, Eq. (3.4) in [7].

$\alpha_s \approx 0.269$ for sequence processing networks [6], and the so-called exactly memorized capacity in [7].

It is found that the iterative form Eq. (24) for the final overlap m_f is effective in most successfully retrieved cases with negligible error. This is evident in Fig. 2, showing that m_f for $N = 100, 200$ until the saturation of the loading ratio α. The iterative results are almost the same as the simulation results when m_f is very close to 1. For example, $\alpha \approx 0.15$ for $\Delta m_f = 0.001$ and $\alpha \approx 0.20$ for $\Delta m_f = 0.002$. When the final overlap parameters m_f deviate from 1, as the loading ratio α increases, the iterative error increases sharply. However, the final overlap parameter has only a small and acceptable error $\Delta m_f < 0.04$ near the saturation $\alpha_s = 0.269$ [see Fig. 2(b)]. As stated above in Eq. (7), we study only the first step behavior of the macroscopic overlap parameters. In order to get more precise results, the signal-to-noise analysis for the first few time steps should be considered [9, 27].

One may be interested in the case where systems can retrieve stored patterns without error. We define this as the precisely retrieved case. In other words, pattern ν at time t and pattern $\nu + 1$ at time $t + 1$ are retrieved without error, which means $m^\nu(t) = 1$, and $m^{\nu+1}(t+1) = 1$. In the stationary state, $m_f = 1$. In our theory, the critical precisely retrieved storage α_c can be calculated from Eq. (24) by setting $m_f = 1$. In Fig. 3, we present α_c obtained by Eq. (24) to compare with the simulation results. Obviously, $\alpha_c \approx 1/\sqrt{N}$ in the iterative algorithm Eq. (24) is consistent with that in the simulation results. Here, it should be mentioned that there is another capacity definition based on absolute stability, $\alpha_A = 1/(2 \log N - \log \log N)$, called the Amari capacity [7]. Obviously, one has the following relationship between the critical loading ratios,

$$\alpha_c < \alpha_A < \alpha_s. \tag{29}$$

Note that the Amari capacity α_A means that there exists some probability for precise retrieval of stored patterns at least one time in a large number of trials. However, the so-called precisely retrieved capacity α_c in this paper means that the systems must precisely retrieve stored patterns for each trial.

B. Random networks

We take more general situations into consideration and explore the effect of degree distributions on the transient dynamics of neural networks in the case of random connection. In the following context, we study three situations: the delta function, binomial, and power-law degree distributions.

In Fig. 4, we plot the temporal evolution of overlaps for the above three types of degree distributions obtained from both the theory [Eq. (27)] and simulations. The parameters are: $N = 50000$, the average degree $\bar{k} = 100$, and the number of stored patterns $p = 20$. The first numerical experiment is the degree distribution with the delta function
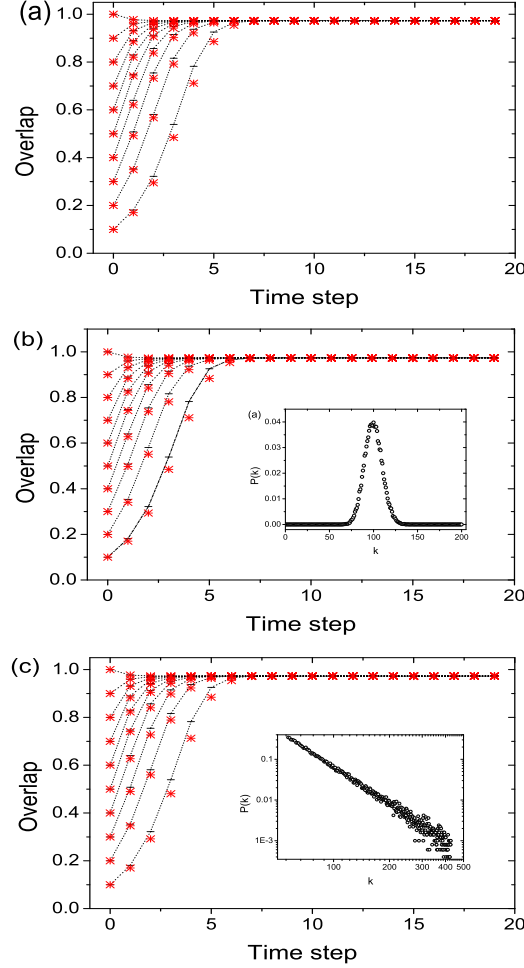
$$P(k) = \delta(k - \bar{k}). \tag{30}$$

FIG. 4: (Color online) The temporal evolution of overlaps from our theory $(-)$ and simulation $(*)$ for (a) delta, (b) binomial, and (c) power-law degree distributions. The parameters of networks are $N = 50000$, $p = 20$, and the average degree $\bar{k} = 100$.

This connection topology is generated by randomizing a regular lattice whose average degree is $\bar{k}$. The temporal evolutions of overlap parameters from the theory and numerical simulations are plotted in Fig. 4(a). The second one is a binomial distribution which comes from an Erdös-Renyi random graph [28] [see the inset of Fig. 4(b)]

$$P(k) = C_N^k \left(\frac{\bar{k}}{N} \right)^k \left(1 - \frac{\bar{k}}{N} \right)^{N-k}. \tag{31}$$

The third one is the power-law distribution [see the inset of Fig. 4(c)],

$$P(k) \sim k^{-3}. \tag{32}$$

The power-law degree distribution can be generated using preferential attachment [29]. It is easy to observe that the theoretical results from our scheme are consistent with the simulations for the three degree distributions above.

Note that the presented cases are all situations of successful retrieval of the stored patterns. Fig. 5 plots the time evolution of overlaps in the case of failed trials with $p = 60$. Apparently, as stated above, encouraging results are also obtained.

Furthermore, we study the effect of size in the network on transient dynamics. Figure 6 shows the comparison between our theory [Eq. (27)] and the numerical simulations for $N = 10000$ and $N = 50000$ in the case of binomial
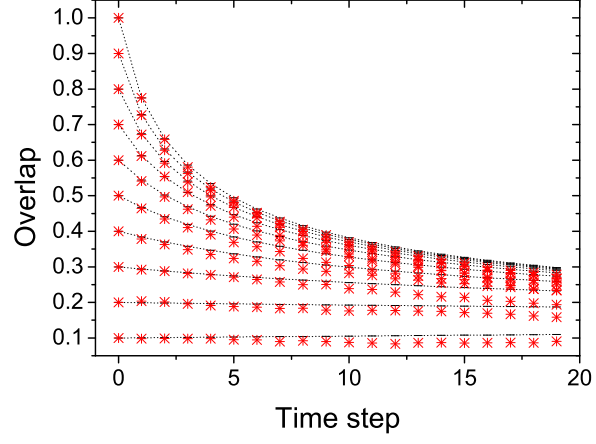
FIG. 5: (Color online) The temporal evolution of overlaps from our theory ($-$) and simulation ($*$) for power-law degree distribution $P(k) \sim k^{-3}$. The parameters of networks are $N = 50000$, $p = 60$, and the average degree $\bar{k} = 100$.
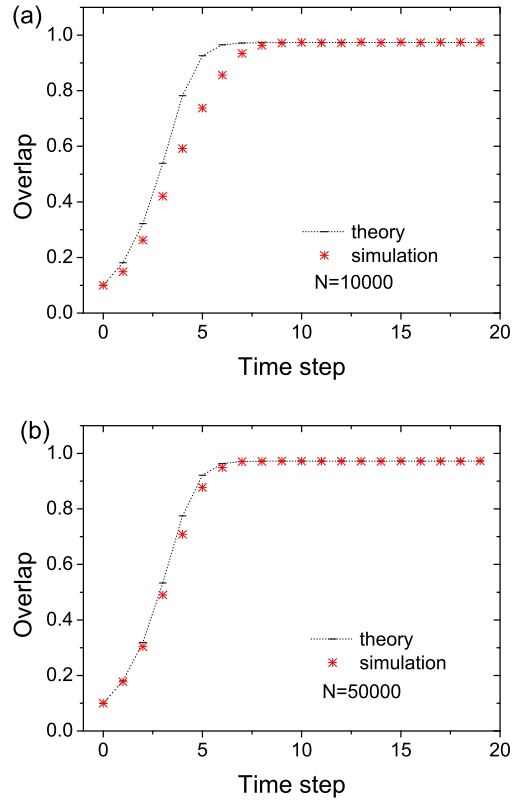

FIG. 6: (Color online) The temporal evolution of overlaps for binomial degree distribution with the average degree $\bar{k} = 100$, $p = 20$, and (a) $N = 10000$, (b) $N = 50000$.

degree distribution. As N increases under all the other same parameters, the theoretical prediction is closer to the simulation result. In fact, this size effect comes from the loop structure in networks. In this paper, our equation (27) does not take into account the loop structure. Loop structure refers to the existence of (perhaps many) short loops in the network, such as triangles or quadrangles (See Ref. [30]). These short loops may cause the coupling of the order parameters at different times and complicate the dynamics. [30] suggested a parameter, loopiness coefficient, to investigate the effect of loop in the networks. Loopiness coefficients grow with $\bar{k}/N$ in random network. Our formula can present better performance for loopiness coefficients smaller with increasing sparseness of network connectivity.

V.   SUMMARY AND CONCLUDING REMARKS

In this paper, we have discussed the effect of degree distribution on transient dynamics for sequence processing neural networks. When the effect of loop structure is absent, we derived the analytic evolution equation for the overlap parameter [Eq. (27)] including the effect of degree distribution that is also obtained in the sparsely connected Hopfield model [26]. In the case of globally coupled networks, the so-called precisely retrieved capacity $\alpha_c = N^{-1/2}$ is suggested by both the theory and simulations; whereas in the case of random networks, our theoretical predictions are consistent with the numerical simulation results under three situations, including the delta, binomial, and power-law degree distributions.

It should be mentioned that, in our presented work, the most efficient arrangement for storage and retrieval of patterns in sequence by the artificial neural network is the random topology. But in real brains, the topology of neural systems appears more complicated and the effect of loop structure becomes inevitable [31, 32]. In a special case, the role of loop structure has been studied without the effect of degree distribution [30]. In future work, we will focus on how to combine the effects from the degree distribution and the loop structure.

Acknowledgment

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175(2006).
[2] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554(1982).
[3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530(1985); Ann. Phys. **173**, 30(1987).
[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007(1985).
[5] A.C.C. Coolen, arXiv:cond-mat/0006011.
[6] A. Düring, A. C. C. Coolen, and D. Sherrington, J. Phys. A **31**, 8607(1998).
[7] S. Amari and K. Maginu, Neural Networks **1**, 63(1988).
[8] M. Okada, Neural Networks **8**, 833(1995).
[9] D. Bollé, J. Busquets Blanco, and T. Verbeiren, J. Phys. A **37**, 1951(2004).
[10] P. N. McGraw and M. Menzinger, Phys. Rev. E **68**, 047102(2003).
[11] L. da Fontoura Costa and D. Stauffer, Physica A **330**, 37(2003).
[12] D. Stauffer, A. Aharony, L. da Fontoura Costa, and J. Adler, Eur. Phys. J. B **38**, 495(2004).
[13] B. J. Kim, Phys. Rev. E **69**, 045101(R)(2004).
[14] J. J. Torres, M. A. Munoz, J. Marro, and P. L. Garrido, Neurocomputing **58**, 229(2004).
[15] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167(1987).
[16] T. L. H. Watkin and D. Sherrington, J. Phys. A **24**, 5427(1991).
[17] P. Zhang and Y. Chen, Lect. Notes Comput. Sci. **4491**, 1144(2007). (cond-mat/0611313).
[18] W. K. Theumann, Physica A **328**, 1(2003).
[19] B. Wemmenhove and A. C. C. Coolen, J. Phys. A **36**, 9617(2003).
[20] H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861(1986).
[21] S. Amari, *Associative Memory and Its Statistical Neurodynamical Analysis*, (Springer, Berlin, 1988).
[22] U. Bauer and U. Krey, Z. Phys. B **79**, 461(1990).
[23] A. C. C. Coolen and D. Sherrington, J. Phys. A **25**, 5493(1992).
[24] O. Melamed, W. Gerstner, W. Maass, M. Tsodyks, and H. Markram, Trends in Neurosciences **27**, 11(2004).
[25] W. Kinzel, Z. Phys. B **60**, 205(1986); **62**, 267(1986).
[26] P. Zhang and Y. Chen, Physica A **387**, 1009(2008).
[27] D. Bollé, G. Jongen, and G. M. Shim, J. Stat. Phys. **91**, 125(1998).

[28] P. Erdös and A. Renyi, Publ. Math. (debrecen) **6**, 290(1959).

[29] A. L. Barabási and R. Albert, Science **286**, 509(1999); A. L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173(1999).

[30] P. Zhang and Y. Chen, cond-mat/0703405, (2007).

[31] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440(1998).

[32] J. W. Scannell, G. Burns, C. C. Hilgetag, M. A. O'Neil, and M. P. Young, Cerebral Cortex **9**, 277(1999).